ELSEVIER

Forensic Science International xxx (2003) xxx–xxx

Forensic Science International

www.elsevier.com/locate/forsciint

High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays

Richard Schoske[a,b], Peter M. Vallone[a], Margaret C. Kline[a], Janette W. Redman[a], John M. Butler[a,*]

[a]*Biotechnology Division, National Institute of Technology, 100 Bureau Drive, Mail Stop 8311, Gaithersburg, MD 20899, USA*
[b]*Department of Chemistry, American University, Washington, DC 20016, USA*

Received 29 April 2003; received in revised form 25 September 2003; accepted 1 October 2003

Abstract

Two Y-chromosome short tandem repeat (STR) multiplex polymerase chain reaction (PCR) assays were used to generate haplotypes for 19 single copy and 3 multi-copy Y-STRs. A total of 27 PCR products were examined in each sample using the following loci: DYS19, DYS385 a/b, DYS388, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS426, DYS437, DYS438, DYS439, DYS447, DYS448, DYS450, DYS456, DYS458, DYS460, DYS464 a/b/c/d, H4, and YCAII a/b. The first multiplex is the Y-STR 20plex previously described by Butler et al. [Forensic Sci. Int. 129 (2002) 10]. The second multiplex is a novel Y-STR 11plex and includes DYS385 a/b, DYS447, DYS448 and the new markers DYS450, DYS456, DYS458, and DYS464 a/b/c/d. These two multiplexes were tested on 647 males from three United States population sample sets: 260 African Americans, 244 Caucasians, and 143 Hispanics. Haplotype comparisons between common loci included in the 20plex and 11plex assays as well as commercially available kits found excellent agreement across a sampling of the population samples. The multi-copy loci DYS464, DYS385, and YCAII were the most polymorphic followed by the following single copy Y-STRs: DYS458, DYS390, DYS447, DYS389II, DYS448, and DYS456. Samples containing the most common type in the European database could be well resolved with additional markers beyond the minimal haplotype loci.

Keywords: Y-chromosome; Y-STR; Multiplex PCR; DNA typing; U.S. populations

1. Introduction

Studying the ability of Y-STR markers to differentiate between DNA samples from unrelated male donors is crucial to the forensic science community, particularly in cases of rape and in paternity testing [1]. The primary drawback of Y-STRs in forensic applications is the absence of recombination between these markers. Thus, while autosomal STR genotypes can differentiate any two individuals with the exception of identical twins, Y-STR haplotypes are less likely to discriminate two unrelated males. There are two ways to increase the potential discrimination capacity of a particular Y-STR haplotype. First, analyze as many Y-STRs as possible. The second and more practical approach is to combine only the most polymorphic markers into a Y-STR multiplex and thereby increase the discrimination capacity of the resulting Y-STR haplotype. Until recently, there were only a limited number of polymorphic Y-STR markers to study. Every year an increased number of markers are being reported in the literature. Recently, 14 novel STRs were published by Redd et al. [2] that have the potential to aid in the discrimination of unrelated males. The discovery of new markers underscores the need for highly multiplexed Y-STR assays to evaluate these markers against previously used loci in the same sample set.

The majority of multiplexes presented in the literature typically involve the simultaneous amplification of 7 or fewer Y-STR markers [3–5]. The European Y chromosome

* Corresponding author. Tel.: +1-301-975-4049; fax: +1-301-975-8505.
E-mail address: john.butler@nist.gov (J.M. Butler).

0379-0738/$ – see front matter © 2003 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2003.10.007

typing community has established a ''minimal haplotype'' and an ''extended haplotype'' for inclusion of common loci into public Y-STR haplotype DNA databases [6,7]. These haplotypes consist of results from the following Y-STR markers: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385, and YCAII.

In this study, the previously described Y-STR 20plex [8] and a novel Y-STR 11plex, which are capable of producing 27 polymorphic Y-chromosome specific PCR amplicons, are used to obtain Y-STR haplotypes on 647 males residing in the United States. The Y-STR 11plex includes the highly polymorphic locus DYS385 a/b, and newly discovered loci DYS447, DYS448, DYS450, DYS456, DYS458, and DYS464 a/b/c/d. A complete characterization of the allelic frequencies of each marker and diversity of various combinations of these markers is presented in an effort to evaluate their utility for the human identity testing community.

2. Materials and methods[1]

2.1. DNA samples

Anonymous liquid blood samples with self-identified ethnicities were purchased from two commercial blood banks: Millennium Biotech, Inc. (Ft. Lauderdale, FL) and Interstate Blood Bank (Memphis, TN). Samples were approved for use through the human subjects internal review board at the National Institute of Standards and Technology (NIST) and were extracted using a modified salt-out procedure [9]. A set of extracted DNA samples from 20 African American and 20 U.S.-Caucasian males were kindly provided by Carll Ladd of the Connecticut Forensic Laboratory (Meriden, CT). Samples from all three sources were tested with 15 autosomal STRs and amelogenin to demonstrate uniqueness and gender [10]. Samples were quantified using UV spectrophotometry and a PicoGreen assay and diluted to approximately 1 ng/μl [10].

2.2. Multiplex primer design and quality control

Consensus sequences for each locus of the Y-STR 11plex were generated using methods previously described for the Y-STR 20plex [8]. The consensus sequence for each of the 22 markers in this study may be found in STRBase [11] http://www.cstl.nist.gov/biotech/strbase/y_strs.htm. The allele ranges of the 22 Y-STR markers listed in Table 1 were defined through analysis of the population samples in this study and do not include all of the alleles identified in the literature. Complete allele ranges for each of these Y-STR markers are available on STRBase [11].

Multiplex PCR design strategy as outlined elsewhere [12] was followed in the development of the Y-STR 11plex and the previously described Y-STR 20plex [8]. Loci for the Y-STR 11plex were organized based on PCR product size range and dye color to accommodate the addition of new markers (see Fig. 1). The Y-STR 11plex was initially designed so the remaining minimal haplotype primer sets shown by Butler et al. [8] (DYS19, DYS389I/II, DYS390, DYS391, DYS392, and DYS393) could potentially be added to the primer mix and thereby increase the number of regions of the Y-chromosome simultaneously amplified to 18.

Primers were then designed with Primer3 [13]. Candidate primers were screened for potential primer cross reactions using an algorithm previously described [14]. Primers were purchased from Qiagen Operon (Alameda, CA) or Applied Biosystems (Foster City, CA) and upon receipt were quality control tested prior to further use to confirm proper synthesis using methods previously described [15]. The primer sequences for the Y-STR 20plex can be found in Butler et al. [8] and those for the Y-STR 11plex are listed in Table 2.

2.3. Y-STR 20plex and Y-STR 11plex PCR amplification conditions

The PCR amplifications for each Y-STR multiplex were performed in reaction volumes of 20 μl with 2 units of AmpliTaq Gold® DNA polymerase (Applied Biosystems), 10 mmol/l Tris–HCl (pH 8.3), 50 mM KCl, 1.75 mmol/l $MgCl_2$, 300 μmol/l deoxynucleotide triphosphates (dNTPs: dATP, dCTP, dGTP, dTTP), 5% glycerol by volume, 0.16μg/μl bovine serum albumin (Sigma, St. Louis, MO), 4.4 μl of the Y-STR primer mix, and 2 μl DNA template. The male DNA template amount in the PCR varied between 1 and 2 ng depending on the volume of the PCR reaction. Some of the PCR reactions were scaled down to 5 or 10 μl. In the case of 5 μl PCR, 1 μl of DNA template was used (approximately 1 ng) and the number of amplification cycles was cut to 26 instead of 28. In the case of 10 μl PCR, 2 μl of DNA template was used (approximately 2 ng) and the number of amplification cycles was cut to 27 instead of 28. Thermal cycling was performed using a GeneAmp 9700 (Applied Biosystems) using the following conditions in 9600-emulation mode (i.e. ramp speeds of 1 °C/s):

95 °C for 10 min
28 cycles:
94 °C for 1 min
55 °C for 1 min
72 °C for 1 min
60 °C for 45 min
25 °C hold

[1] Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology or the U.S. Department of Defense nor does it imply any of the materials, instruments or equipment identified are necessarily the best available for this purpose.

Table 1
Allele ranges and PCR product sizes for 22 Y-STR markers based on population samples examined in this work

STR locus	Allele range	Repeat motif	Size range (bp)
DYS19	12–17	TAGA	241–261
DYS385 a/b—20plex	8–21	GAAA	246–298
DYS385 a/b—11plex	8–21	GAAA	315–367
DYS388	9–17	ATT	148–172
DYS389I	12–15	(TCTG)(TCTA)	155–167
DYS389II	26–34	(TCTG)(TCTA)	263–295
DYS390	20–26	(TCTA)(TCTG)	201–225
DYS391	8–12	TCTA	97–113
DYS392	7–16	TAT	293–320
DYS393	12–16	AGAT	121–133
DYS426	9–13	GTT	89–101
DYS437	13–18	TCTA	182–202
DYS438	8–14	TTTTC	310–340
DYS439	10–15	AGAT	214–234
DYS447—20plex	19–33	TAAWA compound	192–261
DYS447—11plex	19–33	TAAWA compound	192–261
DYS448—20plex	19–27	AGAGAT	293–341
DYS448—11plex	19–27	AGAGAT	330–378
DYS450	6–11	TTTTA	185–210
DYS456	12–18	AGAT	90–114
DYS458	13–20	GAAA	132–160
DYS460	7–12	ATAG	101–121
DYS464 a/b/c/d	11–20	CCTT	250–286
H4	9–14	TAGA	126–146
YCAII a/b	11–25	CA	135–163

Different primers were used for DYS385 and DYS448 between the 20plex and 11plex assays. Multi-copy Y-STR markers are bolded.

2.4. Y-PLEX™ 6 and Y-PLEX™ 5 kits

Y-PLEX™ 6 and Y-PLEX™ 5 kits from ReliaGene Technologies (New Orleans, LA) were used to amplify the Y-STR loci DYS19, DYS385 a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, and DYS439. PCR amplification was performed with 1–2 ng of DNA template according to the manufacturer instructions, except the volumes were reduced to 12.5 or 6.25 μl instead of the suggested 25 μl.

2.5. Detection and analysis of PCR products

The separation and detection of Y-STR 20plex and Y-STR 11plex PCR products was accomplished with the ABI

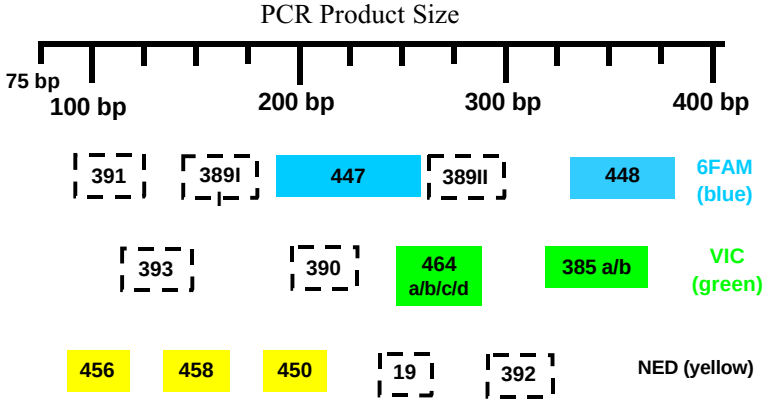

Fig. 1. Schematic of PCR product sizes produced with the allele size ranges for the loci in the Y-STR 11plex. Markers names have been abbreviated (e.g. DYS456 is listed as 456). Markers enclosed with dotted lines indicate minimal haplotype loci (from the 20plex assay [8]) that could be added to Y-STR 11plex without having to redesign any of the primer sets.

Table 2
PCR primer sequences for loci in the Y-STR 11plex

Locus		Primer sequences (5′-to-3′)	Primer3 T_m (°C)
DYS385 a/b	F	**VIC**-TAGACACCATGCCAAACAACA	60.0
	R	GGCTGCTGACCAGATTTCTTT	60.8
DYS447	F	**FAM**-GGTCACAGCATGGCTTGGTT	63.4
	R	GGGCTTGCTTTGCGTTATCTCT	64.0
DYS448	F	**FAM**-GGAGAGGCAAGGATCCAAATA	60.4
	R	GTTGATTCCCTGTGTTGGAGAC	60.0
DYS450	F	**NED**-TGCAGCTGTTTGTAGATCTGGT	60.0
	R	GCCTTTCCAATTTCAATTTCTG	60.0
DYS456	F	**NED**-GGACCTTGTGATAATGTAAGATAG	54.0
	R	GTAGAGGGACAGAACTAATGGAA	55.7
DYS458	F	**NED**-GCAACAGGAATGAAACTCCAAT	60.4
	R	GTTCTGGCATTACAAGCATGAG	59.8
DYS464 a/b/c/d	F	**VIC**-CTTTGGGCTATGCCTCAGTTT	60.6
	R	GCCATACCTGGGTAACAGAGAGAC	59.0

Fluorescent dye labels (6FAM, VIC, or NED) are indicated on the 5′end of the forward primers. Predicted primer melting temperatures (T_m) were calculated as described previously (see [8]). All primer concentrations were 0.4 μmol/l in 20 μl PCR volumes.

Prism® 3100 Genetic Analyzer 16-capillary array system (Applied Biosystems) following manufacturer's protocols using the G5 matrix filter set to detect the five dyes 6FAM™ (blue), VIC™ (green), NED™ (yellow), PET™ (red), and LIZ™ (orange). Prior to sample analysis, a spectral matrix was established using matrix standard set DS-33 (Applied Biosystems). Samples were prepared with 18.6 μl Hi-Di™ formamide (Applied Biosystems), 0.4 μl GS500 LIZ size standard (Applied Biosystems), and with 1 μl of PCR product. Samples were injected 16 at a time for 10 s at 3000 V and separated at 15,000 V for 44 min at a run temperature of 60 °C. Separations were performed using 3700 POP™-6 sieving polymer matrix (Applied Biosystems), 1× A.C.E.™ buffer (Ameresco, Solon, OH), and a 36 cm × 50 μm capillary array (Applied Biosystems). Following data collection, samples were analyzed with Genescan® 3.7 (Applied Biosystems) and allele designations were made in Genotyper 3.7 (Applied Biosystems). Allele calls were assigned based on sizing bin windows of ±0.5 bp rather than by comparison to allelic ladders. The category size of each allele was established using methods fully described elsewhere [16]. Briefly, the sizes of category windows were established through running a reference standard that had been sequenced.

The separation and detection of the Y-PLEX™ 6 and Y-PLEX™ 5 kit generated PCR products were accomplished with an ABI Prism® 310 Genetic Analyzer (Applied Biosystems) using filter set A. The matrix was established with matrix standards for the four dyes FAM (blue), HEX (green), TAMRA (yellow), and ROX (red) (Applied Biosystems). Each sample was prepared by adding 1 μl PCR product to 20 μl of deionized formamide containing 0.75 μl GS500 ROX size standard. Samples were injected for 5 s at 15,000 V and separated at 15,000 V for 26 min with a run temperature of 60 °C using POP™-4 sieving polymer matrix (Applied Biosystems), 1× Genetic Analyzer Buffer with EDTA (Applied Biosystems), and a 47 cm × 50 μm capillary (Applied Biosystems). Following data collection, samples were analyzed with Genescan® 3.1 for Macintosh (Applied Biosystems) and allele calls were determined by comparison to allelic ladders using Genotyper® 2.5 (Applied Biosystems) and the Y-PLEX™ 6 310 v3.0 and Y-PLEX™ 5 310 v1.0 Genotyping template provided by ReliaGene.

2.6. Y-STR nomenclature

The nomenclature for each Y-STR marker is given in Table 1. The repeat motifs for each of the markers were determined using guidance provided by the International Society for Forensic Genetics (ISFG) [17]. It should be noted however that the allele nomenclature for three of the Y-STR markers has been presented inconsistently in the literature. This may impact future databasing efforts if consistent nomenclatures are not followed. These markers are GATA H4, DYS439, and DYS448.

The repeat of Y GATA H4 shown in Table 1 follows the nomenclature of Butler et al. [8] and is different than previously published information [18,19] (see [16] for full details). The reverse primer of the H4 primer set binds internal to some of the invariant repeats and in this study only the variant portion of the repeat is used to designate the allele [8,12]. The nomenclature of DYS439 used in this study only includes the variant portion of the repeat, and excludes the invariant portion as recommended by Gusmao

et al. [19]. Finally, the nomenclature for DYS448 presented here includes a block of 4 invariant repeats previously excluded by Redd et al. [2]. A detailed account of the repeat nomenclature issues for each Y-STR marker is described elsewhere [16].

2.7. Haplotyping of DYS464 a/b/c/d

A single set of DYS464 primers target four separate regions of the Y-chromosome [2]. Thus, up to four polymorphic peaks can be generated per PCR amplification and resolved from one another with electrophoretic separation. There are two potential ways for typing DYS464. Allele designations for DYS464 may be based on the presence or absence of peaks of a different base pair size or as previously noted by Redd et al. [2], allele calls may be assigned on the basis of different peak heights. Fig. 2 illustrates these different approaches to assigning allele calls at the DYS464 locus. The conservative approach is taken here by only using the allele calls without relative peak height information.

2.8. Statistical analysis

Allelic frequencies for each single copy Y-STR locus were calculated by simple counting of observed typing results. The multi-copy loci DYS385 a/b, YCAII a/b, and DYS464 a/b/c/d represent variation at multiple regions of the Y chromosome that cannot be differentiated with the primer pairs used. These allelic patterns are therefore analyzed as ''phenotypes''.

STR diversity (D), or the probability that two alleles chosen at random are different, was calculated using the formula $D = (n/n-1)(1-\sum p_i^2)$, where n is the sample size and p_i is the allelic frequency [20]. Haplotype diversity was computed with the same equation using haplotype frequencies instead of allele frequencies. The probability of obtaining an identical haplotype in a pair of random unrelated males can be estimated as 1-D. Discriminatory capacity was determined by dividing the number of different haplotypes seen in a given population by the total number of samples in that population [2,26]. A unique haplotype is defined as one that occurs only once in a given population.

3. Results and discussion

3.1. Concordance of common loci between Y-STR 20plex and Y-STR 11plex assays

Fig. 3 shows the GeneScan result obtained for the same male DNA sample using the Y-STR 20plex and the Y-STR 11plex assays. Complete Y-STR 20plex and Y-STR 11plex haplotypes were consistently obtained using 1–2 ng of DNA when operating in a high-throughput mode. With these two megaplex assays, Y-STR profiles from 27 PCR products across 647 different male samples can be generated in as little as 1294 PCR reactions. This translates to almost 14, 96-well plates that can be run in less than 100 h operating time on a single ABI 3100 instrument. In short, if a laboratory routinely operated in a high-throughput fashion, the 17,388 allele calls analyzed in this study could have been generated in less than one week by a single instrument and operator.

The Y-STR 11plex and Y-STR 20plex have four overlapping Y-STR markers. Allele calls from these common loci—DYS385 a/b, DYS447, and DYS448—can be compared as a quality assurance measure when running both assays. However, only DYS447 has the same primer

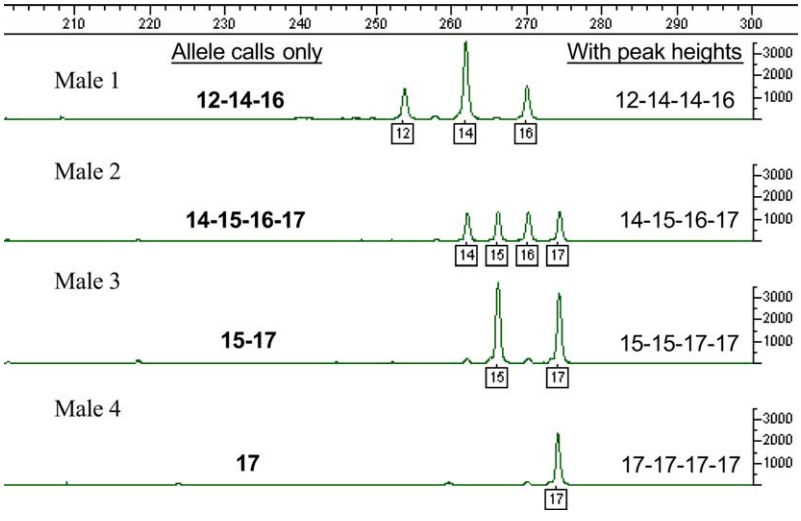

Fig. 2. Genotyper results for the multi-copy Y-STR locus DYS464 illustrating the possibility of 1–4 resolvable PCR products with a single primer pair. The allele calls for the DYS464 amplicons in this work are conservative and are based on the presence of a particular peak rather than considering relative peak heights (see [2]).

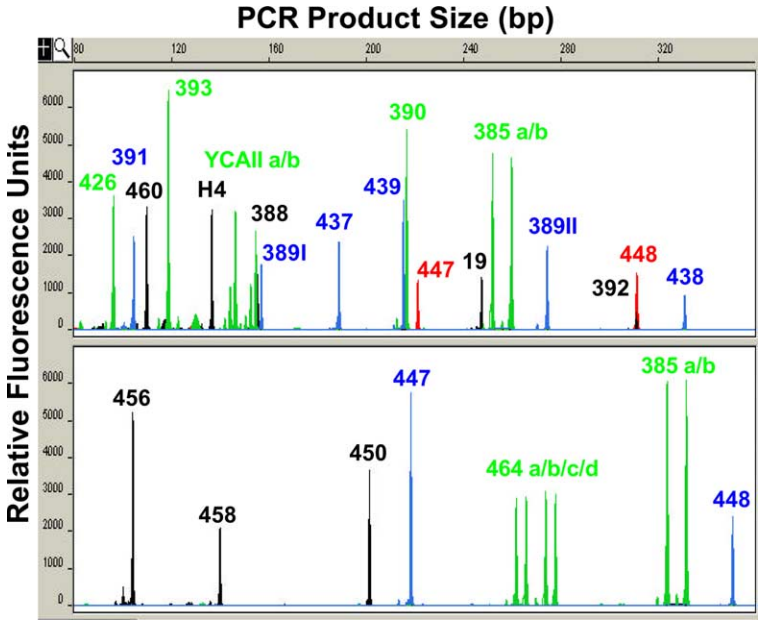

Fig. 3. GeneScan® result obtained from the same male DNA sample amplified with the Y-STR 20plex (top) compared to the Y-STR 11plex (bottom). Each multiplex was amplified using the PCR conditions outlined in the materials and methods with 20 μl reactions, 28 cycles, and 2 ng of DNA template.

sequences between the two multiplexes albeit with different dye labels. The DYS385 a/b and DYS448 PCR products are larger in the Y-STR 11plex assay by 31 and 37 bp, respectively.

The allele calls for both the Y-STR 20plex and the Y-STR 11plex were made with two Genotyper macros that use categories with allele bin windows of ±0.5 bp. The bin windows were established based on sequence information from alleles in NIST Standard Reference Material® SRM 2395 and the allele sizing data provided by an internal sizing standard [16]. A comparison of Y chromosome haplotypes for the 647 male samples used in the population study found excellent agreement out of 2588 possible allele designations between the two multiplexes with the exception of seven calls for the DYS385 a/b locus. The alleles that were the subject of the discrepant calls were identified as microvariants by the Y-STR 11plex assay while the Y-STR 20plex assay identified these same alleles as full repeats. Butler et al. [8] noted this same minor discrepancy in a concordance study comparing the Y-STR 20plex typing information to the Y-PLEX™ 6 kit.

The seven discrepant samples were run using the Y-PLEX™ 6 kit in order to ascertain the correct repeat designation. Each sample was identified as a microvariant instead of the full repeat (see Fig. 4). Amplicon sizes for the DYS385 a/b locus generated by the Y-STR 20plex are smaller than those generated by either the Y-STR 11plex or the Y-PLEX™ 6 kit. Primer positions for the various DYS385 primers in the 20plex and 11plex assays are shown in Fig. 5 in relationship to sequence alignments of a 16.3 and a 17 allele. Both of the top strands for these two DYS385 alleles have 17 GAAA repeats. However, a thymidine deletion is present in a T_7 stretch of the sequence that is *outside* the Y-STR 20plex reverse primer binding site yet *internal* to the Y-STR 11plex reverse primer. The identification of such a deletion to create x.3 alleles was first observed by Furedi et al. [21] and noted in our previous work [8].

3.2. Concordance of Y-STR assays to commercially available test kits

A 40 sample subset of the population samples was run using the Y-PLEX™ 6 and Y-PLEX™ 5 test kits available from ReliaGene Technologies. The resulting haplotypes were compared to those obtained using the Y-STR 20plex and Y-STR 11plex. The two Y-PLEX™ kits generate 11 typing results from 10 different Y-STR markers. The markers for the Y-PLEX 6 kit are: DYS19, DYS389II, DYS390, DYS391, DYS393, and DYS385 a/b; and the Y-PLEX 5 kit includes DYS389I/II, DYS392, DYS438, and DYS439. Each of these markers is contained within the Y-STR 20plex, while the only locus present in the Y-STR 11plex common to either commercial test kit is DYS385 a/b.

A comparison of the Y chromosome haplotypes in the 40 male sample subset indicated that out of 440 possible common allele designations there was complete concordance between the Y-STR 20plex, Y-STR 11plex and the commercially available test kits from ReliaGene with the exception of a single Caucasian sample. This sample was scored as a DYS385 allele 16.3 (with the 11plex and Y-

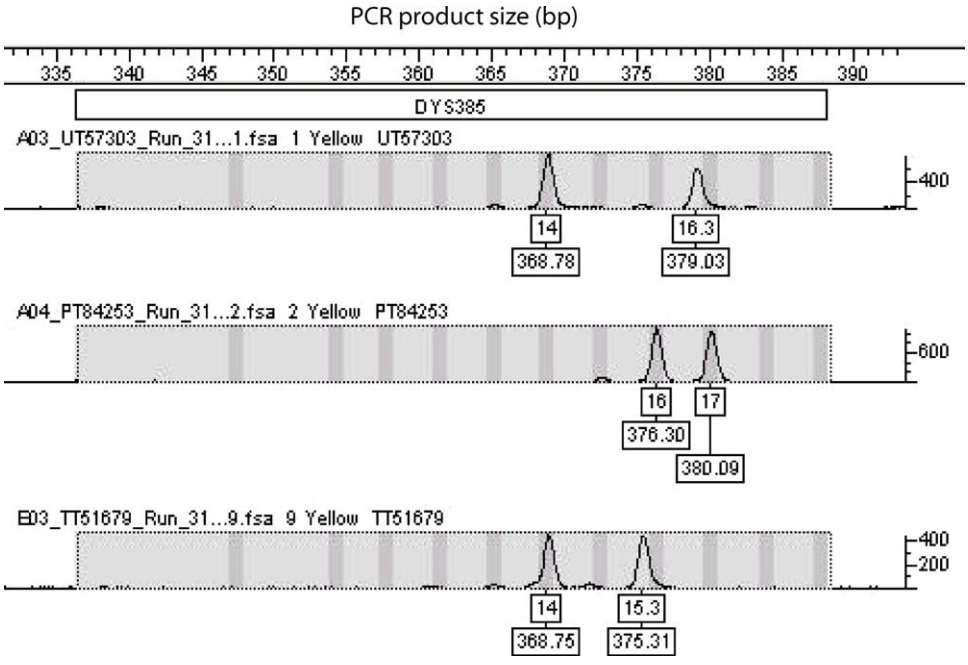

Fig. 4. An example of off-ladder DYS385 alleles 16.3 (top panel) and 15.3 (bottom panel) confirmed through size analysis using the Y-PLEX^{TM} 6 kit. A sample containing alleles 16 and 17 is shown in the center panel for comparison purposes. Peaks are labeled with repeat numbers and calculated size in base pairs (bp).

PLEX^{TM} 6) and an allele 17 (with the 20plex) for the reasons noted above. Due to primer position differences (see Fig. 5), both allele calls could be considered accurate for their particular measurement. It is worth noting that allele calls for the ReliaGene test kits and the in-house assays were generated independently by different personnel using separate instruments platforms, and unique primers with different allele calling approaches. PCR amplicons generated by the Y-PLEX^{TM} 6 and 5 kits were run on an ABI 310 with allele designations being made through comparison of allele sizes to allelic ladders provided in each respective test kit. Alternatively, the Y-STR 20plex and Y-STR 11plex amplicons were run on an ABI 3100 and the allele designations were made using allele bin windows as described above.

This concordance study was designed to mimic inter-laboratory testing procedures performed using Y-STR assays [22–24]. These studies are performed in order to show that for identical DNA samples, numerous laboratories using a variety of separate instrumentation and detection methods can obtain the same typing results. Our mini-study showed that concordance was obtained even though different instrumentation (ABI 310 versus ABI 3100), analysis methods (allelic ladder versus precise sizing method) and different personnel were used to obtain the results.

3.3. Relative diversity values for each Y-STR marker

The STR diversity values for each of the 22 Y-STR markers in each population are shown in Table 3. More polymorphic markers exhibit a higher STR diversity value. All of the markers were ranked according to their STR diversity value by each individual population and then as the total or ''pooled'' population. Although it is not generally recommended to pool allelic frequencies from different ethnic groups (i.e. African Americans and Caucasians) due to statistically significant differences that can exist between these ethnic groups [7,25], a pooled ranking is presented to illustrate the point that even though a marker may not be particularly diverse for a given population it may be valuable in differentiating between samples of different ethnicities.

An example of different allelic frequency distributions that could exist between ethnic groups was observed when the data for DYS390 was examined. DYS390 is one of the markers currently within the minimal haplotype and is within the Y-STR 20plex and the Y-PLEX^{TM} 6 kit. It was the only marker studied that was not one of the 10 most polymorphic markers for each population, but moved into the top 10 when the allelic frequency data was pooled. When evaluated separately DYS390 ranked 8th, 6th, and 13th for the African American, Caucasian, and Hispanic population, respectively. This ranking jumped to 5th when the allelic frequencies were pooled (Table 3). The most common allele for DYS390 in African Americans contained 21 repeats. While the most frequent DYS390 allele for the Caucasians and Hispanics possessed 24 repeat units. Likewise, DYS438 appears to have potential in obtaining different allele frequencies across different population sources (data not shown).

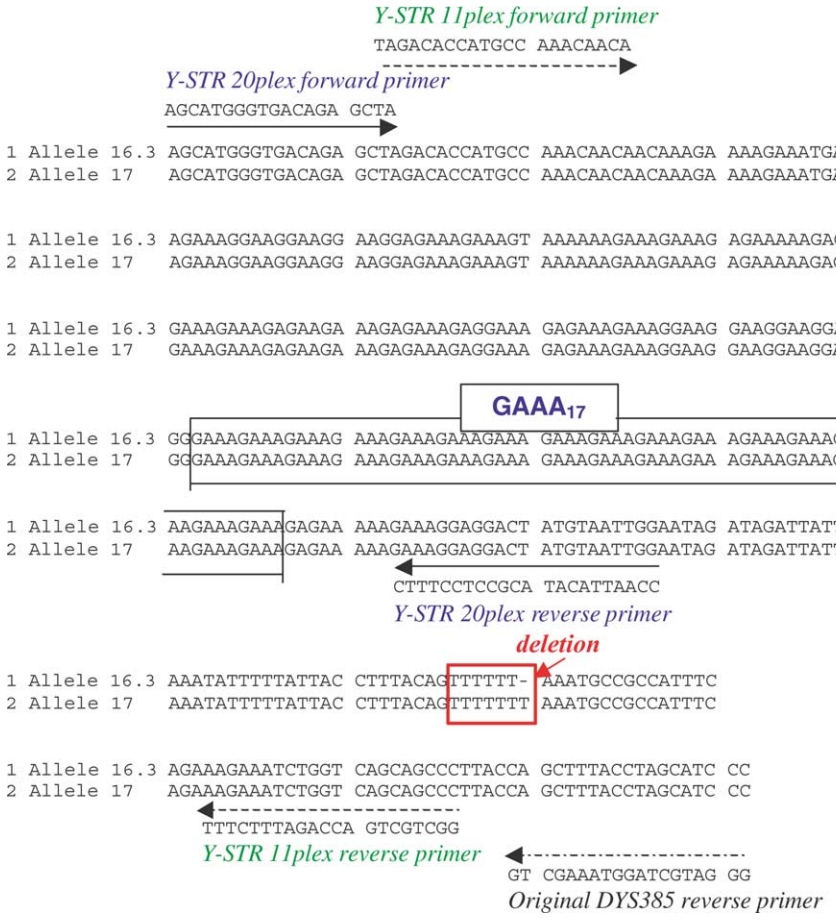

Fig. 5. Alignment of top strands from two different male samples for the DYS385 locus. Both samples contain 17 GAAA tetranucleotide repeats yet one is typed as a ''17'' allele and the other as a ''16.3'' with primers that encompass a polyT stretch (boxed region) 73 bases downstream of the repeat region (see [21]). DYS385 primers internal to the polyT stretch (e.g. 20plex primers) will not be able to measure the deletion that gives rise to this microvariant allele.

Full allelic frequency distributions for loci DYS19, DYS388, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS447, DYS448, DYS450, DYS456, DYS458, DYS460, and H4 have been published elsewhere with this sample set [16]. This data along with the phenotype frequencies for the multi-copy loci DYS385 a/b, YCAII a/b, and DYS464 a/b/c/d are also available on our STRBase website: http://www.cstl.nist.gov/biotech/strbase/NIST-pop.htm. These distributions contain allelic and/or phenotype frequencies for Y-STR markers for three United States ethnic groups: African Americans, Caucasians, and Hispanics. The number of alleles found at each single copy Y-STR marker ranged from 5 to 11 across these samples. For the dual copy loci DYS385 a/b and YCAII a/b, a total of 56 and 27 phenotypes were observed in the three U.S. population sample sets examined. The quadruplicated locus DYS464 a/b/c/d had 110 different phenotypes (Table 4).

The STR diversity values for each Y-STR marker also indicated that even if a marker had a sizeable allele range it might not be significantly polymorphic. The Y-STR DYS392 marker provides a good example of this phenomenon. The allele range obtained in this population study for DYS392 was 7–16 (see Table 1). However, for African Americans DYS392 is one of the least polymorphic, with a STR diversity value of 0.434. This was due to the fact that a majority of the alleles for DYS392 in our African American samples are in the range of 11–13 with over 70% of the alleles having the value of 11. Thus, markers with both large allele ranges and well-balanced allele frequencies are the most polymorphic and usually considered the most desirable in human identity applications. Unfortunately, even though a Y-STR marker may be polymorphic for one ethnic group it may be less variable in another. For example, DYS392 was ranked as the 20th most polymorphic marker in African Americans but ranked 10th and 11th in the Caucasian and Hispanic populations, respectively (Table 3). This STR diversity disparity was also seen in the case of DYS393. In the African American population examined, DYS393

Table 3
STR diversity values among 22 Y-STRs in three U.S. population sample sets

Y-STR	Pooled population		African American		Caucasian		Hispanic	
	STR diversity ($N = 647$)	Rank	STR diversity ($N = 260$)	Rank	STR diversity ($N = 244$)	Rank	STR diversity ($N = 143$)	Rank
DYS464 a/b/c/d	0.956	1	0.954	1	0.934	1	0.937	1
DYS385 a/b	0.912	2	0.942	2	0.838	2	0.901	2
YCAII a/b	0.790	3	0.797	3	0.701	5	0.772	4
DYS458	0.765	4	0.758	5	0.743	3	0.793	3
DYS390	0.764	5	0.664	10	0.701	5	0.665	13
DYS447	0.747	6	0.767	4	0.683	7	0.748	5
DYS389II	0.736	7	0.722	6	0.675	8	0.734	6
DYS448	0.721	8	0.722	6	0.595	11	0.704	8
DYS456	0.700	9	0.671	9	0.731	4	0.695	9
DYS438	0.691	10	0.560	15	0.594	12	0.690	10
DYS19	0.676	11	0.722	6	0.498	19	0.672	12
DYS439	0.656	12	0.636	11	0.639	9	0.717	7
DYS437	0.637	13	0.499	17	0.583	13	0.624	14
H4	0.611	14	0.612	12	0.562	14	0.609	15
DYS392	0.609	15	0.434	20	0.596	10	0.673	11
DYS460	0.570	16	0.568	14	0.555	15	0.556	18
DYS389I	0.549	17	0.531	16	0.538	17	0.596	16
DYS391	0.534	18	0.447	19	0.552	16	0.577	17
DYS426	0.519	19	0.375	21	0.482	20	0.522	19
DYS450	0.489	20	0.487	18	0.177	22	0.414	21
DYS393	0.485	21	0.586	13	0.363	21	0.448	20
DYS388	0.365	22	0.246	22	0.501	18	0.312	22

The multi-copy Y-STRs DYS464, DYS385, and YCAII are the most diverse across all three populations. Only two of the minimal haplotype loci (bolded) occur in the top 10 of the markers tested across all populations. The rank order for the Y-STRs in the pooled population is maintained in the individual population listings.

ranked 13th with a STR diversity value 0.586, but dropped to a ranking of 21st and 20th in the Caucasian and Hispanic populations, respectively.

The most polymorphic markers (i.e. containing the highest STR diversity values) across the population samples we examined were the multi-copy loci DYS385, YCAII, and DYS464. The STR diversity values ranged from a low of 0.177 for DYS450 in Caucasians to a high of 0.954 for DYS464 in African Americans (Table 3). Of the top 10 Y-STR markers according to STR diversity values in each ethnic group, 7 were common to each population. These seven markers included DYS464 a/b/c/d, DYS385 a/b, YCAII a/b, DYS447, DYS456, DYS458, and DYS389II. Only two of these Y-STRs (DYS385 a/b and DYS389II) are found in the European ''minimal haplotype'' and present in current commercially available test kits.

3.4. Haplotype diversity

Once the complete data set had been generated, various combinations of Y-STR markers were examined to gain insight on relative performance of these markers and haplotypes. The ''complete'' Y-STR haplotypes in this study, consisting of 27 allele calls per sample, for each U.S. population (African American, Caucasian, and Hispanic) have been described elsewhere [16] and will be made available on STRBase [11]. The most common way to measure the ability of a Y-STR assay to resolve two unrelated male samples is to calculate its haplotype diversity value and corresponding random match probability [20].

The various combinations of markers studied include the Y-PLEX 6 kit, the widely used ''minimal'' and ''extended'' haplotypes [6], the ''U.S. haplotype'', and the Y-STR 20plex and Y-STR 11plex assays separately and together. The haplotype diversity value and random match probability for each of the various combinations of markers is given in Table 5.

Another purpose of examining combinations of Y-STR markers is to see if viable substitutes can be found for the highly polymorphic multi-copy dinucleotide and stutter-prone YCAII. YCAII was first included in the extended haplotype because it increased the number of unique haplotypes in European population data sets [6]. However, YCAII is not an ideal marker when studying samples involving male:male mixtures due to a high degree of stutter [2].

In January 2003, the Scientific Working Group on DNA Analysis Methods (SWGDAM) selected the initial core Y-STR loci to be used in the United States. This ''U.S. haplotype'' includes the widely used minimal haplotype loci plus DYS438 and DYS439 and can be obtained by utilizing the Y-PLEX™ 5 and Y-PLEX™ 6 kits from

Table 4
Phenotype frequencies for three U.S. ethnic groups using the multi-copy Y-STR locus DYS464 (see [2]) and a conservative typing approach (see Fig. 2)

Locus	Phenotype	African Americans ($N = 260$)	Caucasians ($N = 244$)	Hispanics ($N = 143$)	Pooled ($N = 647$)
DYS464	14		0.012		0.005
	15		0.004		0.002
	16		0.004		0.002
	17		0.004		0.002
	11-12-14-15			0.007	0.002
	11-13-14-15		0.004	0.007	0.003
	11-13-15-16			0.007	0.002
	11-13-**15.1**-16			0.007	0.002
	11-13-16			0.007	0.002
	11-14			0.014	0.003
	11-14-15	0.004	0.016		0.009
	11-14-16	0.004	0.012		0.006
	11-15-16			0.007	0.002
	11-15-16-17	0.004			0.002
	11-15-18		0.004		0.002
	11-16	0.004			0.002
	11-16-18	0.008			0.003
	12-13-14	0.004	0.025	0.014	0.014
	12-13-14-15	0.004	0.004		0.003
	12-13-14-16		0.008	0.007	0.005
	12-13-15-17	0.004			0.002
	12-14		0.004	0.007	0.003
	12-**14.3**			0.007	0.002
	12-14-15	0.008	0.033	0.007	0.017
	12-14-15-16	0.012	0.061	0.021	0.032
	12-14-15-17	0.004			0.002
	12-14-16		0.008	0.007	0.005
	12-14-16-18	0.008			0.003
	12-14-17		0.008	0.007	0.005
	12-15	0.008	0.008		0.006
	12-15-16	0.004	0.045		0.019
	12-15-16-17	0.012	0.008		0.008
	12-15-16-18	0.023			0.009
	12-15-17	0.004		0.007	0.003
	12-15-17-18	0.004			0.002
DYS464	12-15-17-19	0.004			0.002
	12-15-18	0.004			0.002
	12-16	0.008		0.007	0.005
	12-16-17	0.012			0.005
	12-16-19			0.014	0.003
	12-16-20	0.004			0.002
	13-**13.1**-15-16	0.004	0.004		0.003
	13-14	0.004	0.012	0.014	0.009
	13-14-15		0.008		0.003
	13-14-15-16	0.008			0.003
	13-14-15-17	0.004			0.002
	13-14-15-18			0.007	0.002
	13-14-16	0.004			0.002
	13-14-16-17	0.004			0.002
	13-14-16-18			0.007	0.002
	13-14-17	0.004		0.007	0.003
	13-**14.3**-15-17			0.007	0.002
	13-**14.3**-16-17	0.004			0.002
	13-15	0.004	0.004	0.007	0.005
	13-15-**15.1**		0.004		0.002

Table 4 (*Continued*)

Locus	Phenotype	African Americans ($N = 260$)	Caucasians ($N = 244$)	Hispanics ($N = 143$)	Pooled ($N = 647$)
	13-15-16	0.042	0.004	0.014	0.022
	13-15-16-17	0.019	0.008	0.007	0.012
	13-15-16-18	0.038			0.015
	13-15-17	0.004	0.004	0.007	0.005
	13-15-17-18	0.012			0.005
	13-15-18	0.012	0.004	0.007	0.008
	13-15-19		0.008		0.003
	13-**15.3**-17		0.004		0.002
	13-16		0.004	0.014	0.012
	13-16-17	0.146		0.014	0.062
	13-16-17-18	0.031			0.012
	13-16-17-19	0.004			0.002
	13-16-18	0.092	0.004	0.021	0.043
	13-16-18-19	0.004	0.004		0.003
	13-16-19	0.015		0.007	0.008
	13-17-18	0.008			0.003
	13-17-20			0.007	0.002
	14-**14.3**-16-18	0.004			0.002
	14-**14.3**		0.004		0.002
DYS464	**14.3**-15-16-17	0.004			0.002
	14-15	0.004	0.025	0.021	0.015
	14-15-16	0.012	0.016	0.021	0.015
	14-15-16-17	0.004	0.016	0.021	0.012
	14-15-16-18	0.012		0.014	0.008
	14-15-17	0.019	0.053	0.021	0.032
	14-15-17-18	0.012	0.016	0.014	0.014
	14-15-18	0.004	0.004	0.014	0.006
	14-**15.3**-17		0.008	0.007	0.005
	14-16			0.021	0.005
	14-16-17	0.019	0.004	0.056	0.022
	14-16-17-18	0.015	0.004		0.008
	14-16-18	0.008			0.003
	14-16-20	0.004			0.002
	14-17		0.008	0.021	0.008
	14-17-18-19	0.004			0.002
	15-**15.3**-16		0.004		0.002
	15-**15.3**-17-18			0.007	0.002
	15-16	0.042	0.029	0.049	0.039
	15-16-17	0.077	0.152	0.147	0.121
	15-16-17-18	0.004	0.025	0.007	0.012
	15-16-17-19		0.004		0.002
	15-16-18	0.008	0.029	0.014	0.017
	15-16-18-20		0.004		0.002
	15-16-19		0.004		0.002
	15-16-20		0.004		0.002
	15-17	0.050	0.172	0.189	0.127
	15-17-18	0.027	0.037	0.021	0.029
	15-17-18-19		0.004		0.002
	15-17-19		0.004		0.002
	15-18	0.004	0.004	0.014	0.006
	15-18-19	0.004		0.007	0.003
	16-17	0.019	0.008		0.011
	16-18		0.004		0.002
	16-19	0.004			0.002
	17-18	0.004			0.002

Microvariant alleles are bolded.

Table 5
Haplotype diversity (HD) and random match probability (RMP) for various Y-STR marker combinations

Y-STR marker combinations	African Americans ($N = 260$)		Caucasians ($N = 244$)		Hispanics ($N = 143$)	
	HD	RMP	HD	RMP	HD	RMP
Y-PLEX 6 kit	0.9974	0.0026	0.9914	0.0086	0.9934	0.0066
''Minimal'' haplotype	0.9982	0.0018	0.9946	0.0053	0.9957	0.0043
''Extended'' haplotype	0.9988	0.0012	0.9971	0.0029	0.9975	0.0025
''U.S. haplotype''	0.9993	0.0007	0.9974	0.0026	0.9986	0.0014
Y-STR 11plex	0.9993	0.0007	0.9987	0.0013	0.9992	0.0008
Y-STR 20plex	0.9998	0.0002	0.9998	0.0002	0.9998	0.0002
22 Y-STRs	0.9999	0.0001	0.9999	0.0001	0.9999	0.0001
Top 10 (without YCAII a/b)	0.9999	0.0001	0.9999	0.0001	0.9999	0.0001

The ''U.S. haplotype'' includes the minimal haplotype loci plus DYS438 and DYS439. The ''top 10 loci (without YCAII a/b)'' used for this analysis are DYS464 a/b/c/d, DYS385 a/b, DYS458, DYS390, DYS447, DYS389II, DYS448, DYS456, DYS438, and DYS389I.

ReliaGene. Most forensic laboratories in the United States do not use ''home-brew'' assays. Therefore, these labs only use and validate commercially available kits. This study is the first attempt to compare the new ''U.S. haplotype'' with the ''extended'' haplotype and determine if the combination of DYS438 and DYS439 are a suitable replacement for YCAII.

Finally, the combinations of markers in the Y-STR 20plex, Y-STR 11plex, and the most polymorphic markers according to their STR diversity values were examined to determine if any of the other markers not currently within the United States haplotype would make a potential Y-STR DNA profile more discriminatory.

The Y-PLEX^{TM} 6 kit had the smallest haplotype diversity (HD) value for all three ethnic groups because it possessed the fewest markers (the Y-PLEX^{TM} 5 kit was not examined by itself because it does not contain DYS385). The Y-PLEX^{TM} 6 HD in African Americans was 0.9974 but only 0.9914 in Caucasians (Table 5). In general, the HD value and random match probability (RMP) of various combinations increased in each population as the number of markers increased. When all 22 markers were included in the calculation of HD, its value reached 0.9999 for all populations with a RMP of only 0.01%. By just looking at the change in HD values, an argument can be made that no significant benefit in discrimination is obtained by increasing the number of markers past those in the extended or U.S. haplotype. For example, in African Americans the HD value for the U.S. haplotype is 0.9933. This is 0.0056 less than the 0.9999 value for the 22 Y-STR markers. However, examining the discriminatory capacity and number of unique haplotypes observed per population illustrates the value of increasing the number of markers analyzed.

3.5. Discriminatory capacity of haplotypes

The discriminatory capacity (DC) and the number of unique haplotypes for each Y-STR marker combination studied are listed in Table 6. The DC is a measure of how many different haplotypes were observed for a given population [26]. For example in the case of Caucasians, there were 168 different Y-PLEX^{TM} 6 kit haplotypes observed. Of those 168 different haplotypes; 136 occurred once, 20 occurred twice, 5 occurred 3 times, 1 occurred 4 times, 2 occurred 5 times, 2 occurred 8 times, 1 occurred 11 times, and 1 occurred 12 times. For Caucasians, the Y-PLEX^{TM} 6 kit markers had a DC of 68.9% (Table 6). Thus, while the Y-PLEX^{TM} 6 kit had a 0.9914 HD value for Caucasians, a significant portion of the population was not resolved using the markers in the Y-PLEX^{TM} 6 kit. When all 22 markers are included in the calculation of DC, the number of different haplotypes moves from 168 to 242 out of a possible 244 (Table 6). All of the Caucasians were distinguishable from one another using the 22-marker haplotype with the exception of two samples. Measuring DC helps show the importance of studying as many markers as possible.

As expected, the greatest diversity and discriminatory capacity of Y-STR haplotypes was seen when all 22 markers were used (Tables 5 and 6). The DC percentage for the 22 Y-STR haplotype was over 98% for each population analyzed in this study. Out of 647 samples examined, six Y-STR haplotypes occurred twice, one profile was present three times and there were 632 unique Y-STR haplotypes. It should be noted that each of these 647 samples were tested with 15 autosomal STRs and shown to be unique [10]. Four of the six duplicates were population specific. The other two pairs matched an African American with a Caucasian, and a Hispanic with a Caucasian. The haplotype that occurred three times involved two Hispanics and one African American.

It is worth investigating if smaller combinations of carefully chosen highly polymorphic markers can produce similar DC values when compared to larger combinations of markers. The DC value was calculated for the top 10 most diverse Y-STR markers minus the stutter prone YCAII across the pooled population. These 10 markers were: DYS464 a/b/c/d, DYS385 a/b, DYS458, DYS390, DYS447,

Table 6
Discriminatory capacity (DC) and number of unique haplotypes (UH) for the three U.S. populations examined in this study

Y-STR marker combinations	African Americans ($N = 260$)		Caucasians ($N = 260$)		Hispanics ($N = 260$)	
	DC (%)	UH	DC (%)	UH	DC (%)	UH
Y-PLEX 6 kit	82.3	188	68.9	136	78.3	97
"Minimal" haplotype	88.5	213	75.8	161	81.1	100
"Extended" haplotype	91.9	227	83.6	184	89.5	120
"U.S. haplotype"	91.9	222	82.3	176	93.3	121
Y-STR 11plex	93.1	227	88.5	198	94.4	127
Y-STR 20plex	98.5	252	97.2	230	98.6	139
22 Y-STR markers	98.9	254	99.6	242	99.3	141
Top 10 (without YCAII a/b)	96.9	244	97.5	232	99.3	141

DC is defined as the number of different haplotypes observed in the population divided by the population size. UH is defined as the number of haplotypes that occur only once in a given population. The "U.S. haplotype" includes the minimal haplotype loci plus DYS438 and DYS439. The "top10 loci (without YCAII a/b)" used for this analysis are DYS464 a/b/c/d, DYS385 a/b, DYS458, DYS390, DYS447, DYS389II, DYS448, DYS456, DYS438, and DYS389I.

DYS389I/II, DYS448, DYS456, and DYS438. For the purposes of this study, these "top ten" markers have not yet been actually combined into a multiplex for testing in the laboratory. The DC value was calculated to illustrate the potential value of constructing such a multiplex. DYS389I was included in this group because the primers used in this study to amplify DYS389II would also amplify DYS389I.

The discriminatory capacity of these markers was similar to the 22-marker Y-STR profile for all three ethnicities (Table 6). In both Caucasian and African American sample sets, there were only 10 more unique haplotypes with 22-marker Y-STR profiles than could be obtained with the top 10 most diverse markers. In Hispanics, there was no difference between the Y-STR 22 marker profile and the profile obtained by using only the most polymorphic markers. Thus, the addition of other markers beyond the most polymorphic ones led to minimal improvement in the discriminatory capacity with these samples.

3.6. Resolving the most common minimal haplotype

Within the European Y-STR Haplotype Database [6], the most frequent minimal haplotype occurs approximately 3% of the time. This most common type (MCT) has the following alleles: DYS19-14, DYS389I-13, DYS389II-29, DYS390-24, DYS391-11, DYS392-13, DYS393-13, and DYS385a/b-11,14. In the only previously published U.S. population study with the minimal haplotype loci (see http://www.ystr.org/usa), this MCT was observed in 9 out of 599 African Americans, 25 out of 628 Caucasians, and 19 out of 478 Hispanic samples [7,25]. The addition of new Y-STR markers in the present study merits an examination of their

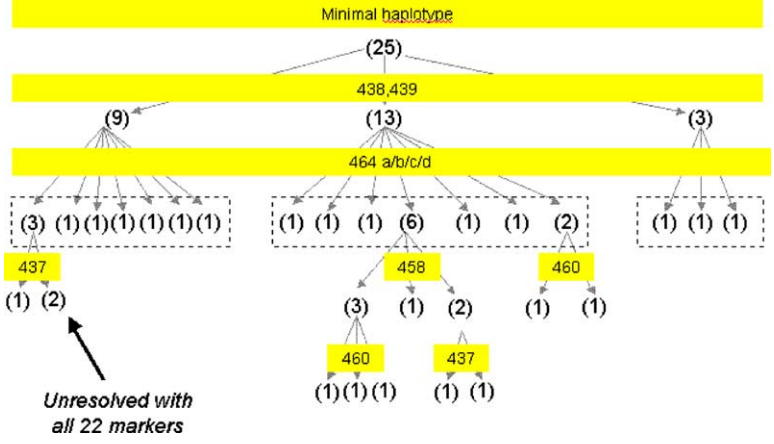

Fig. 6. A total of 25 individuals in the 647 U.S. males tested here possess the most common minimal haplotype (14-13-29-24-11-13-13-11,14), which makes up ~3% of the European Y-STR Haplotype Database. With the addition of DYS438 and DYS439 followed by DYS464, DYS458, DYS460, and DYS437, all but one of the pairs can be resolved from one another down to individual (1) samples.

ability to separate samples that the minimal haplotype cannot. This particular minimal MCT haplotype occurred 25 times in our population set (Fig. 6).

If typing results for DYS438 and DYS439 are added to the MCT haplotype (i.e. the U.S. haplotype markers), the single block of 25 male samples can be resolved down to 3 blocks of 9, 13, and 3 MCT males (Fig. 6). The highly polymorphic DYS464 multi-copy locus can further differentiate the original 25 samples into 17 different groups with all but three of them containing singlets. DYS437, DYS458, and DYS460 help resolve all additional pairs of samples except one (Fig. 6).

4. Implications and conclusions

Y-STR typing data for 22 markers that amplify 27 regions of the Y-chromosome has been presented and analyzed from 647 different male samples in three U.S. populations. The Y-STR markers used in this study have been ranked in terms of their ability to differentiate between unrelated male samples. The discriminatory capacities of various combinations of markers, including those in commercial test kits, have been explored. The ability to obtain information from 27 different regions of the Y chromosome in only two different amplifications should greatly speed up Y-STR database development. For example, with a single ABI 3100 and these two megaplex amplifications, we were able to generate U.S. population data roughly the size of previously available databases [5,7] with only one week of instrument operating time (less than 100 h).

By performing concordance studies, which relied on the use of different primer sets for identical loci, we were able to identify microvariant alleles for the DYS385 a/b locus. These microvariants were not detected using the Y-STR 20plex but were identified when the DYS385 a/b primer sets in the Y-STR 11plex and Y-PLEX^{TM} 6 kit were used. Future databasing efforts of DYS385 a/b should include a note of whether or not the deletion region in the 3′-flanking region was encompassed by the primer set (see Fig. 5). Thus, if one lab obtains a 16.3 variant allele for a sample and a different lab obtains a 17, the difference can be attributed to the different primer sets used and not laboratory error.

The haplotype data described here corroborates with what was described by Redd et al. [2] in demonstrating that there are Y-STR markers more diverse than the majority of markers in the ‘‘minimal’’ haplotype. It would be interesting to see if the non-minimal haplotype markers in this work could resolve any of the matches found in the http://www.y-str.org/usa database [7,25]. DYS464 a/b/c/d, DYS456, and DYS458 that were all included in the Y-STR 11plex, have STR diversity values of 0.67 or greater (Table 3). Unfortunately, none of these markers are widely used yet and are not in any of the currently available commercial test kits. Furthermore, by using just the most polymorphic markers, a high discrimination capacity can be obtained and thus limit the need for testing of additional markers.

The U.S. haplotype (minimal haplotype + DYS438, DYS439) appears to offer a viable alternative to the European ‘‘extended’’ haplotype. In each population, the DC for the U.S. haplotype was similar to the extended haplotype value (Table 6). The DYS438 and DYS439 loci should be able to replace YCAII without any appreciable loss in discriminatory capability and without the problematic stutter products seen in YCAII amplicons.

In summary, the high-throughput Y-STR typing of 22 markers using two multiplexes on a single ABI 3100 Genetic Analyzer allowed us to rapidly generate a database on 647 U.S. population samples. This data provides valuable information to companies and/or research laboratories in the forensic community developing Y-STR assays by directing them to which markers offer the highest diversity and thus should be included in future multiplexes.

Acknowledgements

This work is funded by an interagency agreement between the National Institute of Justice and the NIST Office of Law Enforcement Standards. Richard Schoske is a Ph.D. candidate at American University under Professor Jim Girard and is funded by the United States Air Force, through the Air Force Institute of Technology. The authors would like to thank Dr. Alan Redd and Dr. Michael Hammer for early access to loci included in the Y-STR 11plex.

References

[1] M.A. Jobling, A. Pandya, C. Tyler-Smith, Int. J. Legal Med. 110 (1997) 118–124.
[2] A.J. Redd, A.B. Agellon, V.A. Kearney, V.A. Contreras, T. Karafet, H. Park, P. de Knijff, J.M. Butler, M.F. Hammer, Forensic Sci. Int. 130 (2002) 97–111.
[3] A.J. Redd, S.L. Clifford, M. Stoneking, Biol. Chem. 378 (1997) 923–927.
[4] M. Prinz, K. Boll, H. Baum, B. Shaler, Forensic Sci. Int. 85 (1997) 209–218.
[5] S. Sinha, B. Budowle, S. Arcot, S. Bailey, S. Richey, R. Chakraborty, M. Jones, P. Wojtkiewicz, D. Schoenbauer, A.M. Gross, S.K. Sinha, J.G. Shewale, J. Forensic Sci. 48 (2003) 93–103.
[6] L. Roewer, M. Krawczak, S. Willuweit, M. Nagy, C. Alves, A. Amorim, K. Anslinger, C. Augustin, A. Betz, E. Bosch, A. Caglia, A. Carracedo, D. Corach, A. Dekairelle, T. Dobosz, B.M. Dupuy, S. Furedi, C. Gehrig, L. Gusmao, J. Henke, L. Henke, M. Hidding, C. Hohoff, B. Hoste, M.A. Jobling, H.J. Kargel, P. de Knijff, R. Lessig, E. Liebeherr, M. Lorente, B. Martinez-Jarreta, P. Nievas, M. Nowak, W. Parson, V.L. Pascali, G. Penacino, R. Ploski, B. Rolf, A. Sala, U. Schmidt, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, M. Kayser, Forensic Sci. Int. 118 (2001) 106–113.
[7] M. Kayser, S. Brauer, S. Willuweit, H. Schadlich, M.A. Batzer, J. Zawacki, M. Prinz, L. Roewer, M. Stoneking, J. Forensic Sci. 47 (2002) 513–519.
[8] J.M. Butler, R. Schoske, P.M. Vallone, M.C. Kline, A.J. Redd, M.F. Hammer, Forensic Sci. Int. 129 (2002) 10–24.

[9] S.A. Miller, D.D. Dykes, H.F. Polesky, Nucleic Acids Res. 16 (1988) 1215.
[10] J.M. Butler, R. Schoske, P.M. Vallone, J.W. Redman, M.C. Kline, J. Forensic Sci. 48 (2003) 908–911.
[11] C.M. Ruitberg, D.J. Reeder, J.M. Butler, Nucleic Acid Res. 29 (2001) 320–322.
[12] R. Schoske, P.M. Vallone, C.R. Ruitberg, J.M. Butler, Anal. Bioanal. Chem. 375 (2003) 333–343.
[13] S. Rozen, H.J. Skaletsky, Primer3 (primer3_www.cgi v 0.2) code available at http://www-genome-wi.mit.edu/genome_-software/other/primer3.html.
[14] J.M. Butler, C.M. Ruitberg, P.M. Vallone, Fresenius J. Anal. Chem. 369 (2001) 200–205.
[15] J.M. Butler, J.M. Devaney, M.A. Marino, P.M. Vallone, Forensic Sci. Int. 119 (2001) 87–96.
[16] R. Schoske, The design, optimization and testing of Y chromosome short tandem repeat megaplexes, Ph.D. dissertation, American University, 2003 available as a pdf file at http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm.
[17] P. Gill, C. Brenner, B. Brinkmann, B. Budowle, A. Carracedo, M.A. Jobling, P. de Knijff, M. Kayser, M. Krawczak, W.R. Mayr, N. Morling, B. Olaisen, V. Pascali, M. Prinz, L. Roewer, P.M. Schneider, A. Sajantila, C. Tyler-Smith, Int. J. Legal Med. 114 (2001) 104–109.
[18] A. Gonzalez-Neira, M. Elmoznino, M.V. Lareu, P. Sanchez-Diz, L. Gusmao, M. Prinz, A. Carracedo, Forensic Sci. Int. 122 (2001) 19–26.
[19] L. Gusmao, C. Alves, S. Alves, S. Costa, A. Amorim, M. Brion, A. Gonzalez-Neira, P. Sanchez-Diz, A. Carracedo, Int. J. Legal Med. 116 (2002) 322–326.
[20] E. Bosch, A.C. Lee, F. Calafell, E. Arroyo, P. Henneman, P. de Knijff, M. Jobling, Forensic Sci. Int. 125 (2002) 42–51.
[21] S. Furedi, J. Woller, Z. Padar, M. Angyal, Int. J. Legal Med. 113 (1999) 38–42.
[22] A. Carracedo, A. Beckmann, A. Bengs, B. Brinkmann, A. Caglia, C. Capelli, P. Gill, L. Gusmao, C. Hagelberg, C. Hohoff, B. Hoste, A. Kihlgren, A. Kloosterman, B.M. Dupuy, N. Morling, G. O'Donnell, W. Parson, C. Phillips, M. Pouwels, R. Scheithauser, H. Schmitter, P.M. Schneider, J. Schumm, I. Skitsa, B. Strandmann-Bellinghausen, M. Stuart, D.S. Court, C. Vide, Forensic Sci. Int. 119 (2001) 28–41.
[23] S. Presciuttini, A. Caglia, M. Alu, A. Asmundo, L. Buscemi, L. Caenazzo, E. Carnevali, E. Carra, Z. De Battisti, F. De Stefano, R. Domenici, A. Piccinini, N. Resta, U. Ricci, V.L. Pascali, Forensic Sci. Int. 122 (2001) 184–188.
[24] P.M. Schneider, E. d'Aloja, B.M. Dupuy, B. Eriksen, A. Jangblad, A.D. Kloosterman, A. Kratzer, M.V. Lareu, H. Pfitzinger, S. Rand, R. Scheithauer, H. Schmitter, I. Skitsa, D. Syndercombe-Court, M.C. Vide, Forensic Sci. Int. 102 (1999) 159–165.
[25] M. Kayser, S. Brauer, H. Schadlich, M. Prinz, M.A. Batzer, P.A. Zimmerman, B.A. Boatin, M. Stoneking, Genome Res. 13 (2003) 624–634.
[26] M. Kayser, et al. Int. J. Legal Med. 110 (1997) 125–133.